INF 2178-A4
Mingrui Fu 1010506551

Exploring Cognitive Function and Brain Volume in MRI Results

1. Introduction

In this assignment, we embark on an exploration and analysis of MRI data extracted from a subset of a longitudinal study focusing on patients with/without dementia. The dataset titled "INF2178_A4_data.csv" serves as the foundation for our investigation into within-subject design and statistical power using mixed-effects ANOVA. Our analysis is motivated by two distinct research questions:

Research Question #1: What is the impact of gender (M/F) and visit time points (Visit) on cognitive function, as measured by MMSE scores, while considering the interaction between gender and visit time points?

Research Question #2: Is there a significant difference in brain volume (eTIV) across multiple visits for each patient?

Leveraging Python, we will conduct quantitative analysis utilizing mixed-effects ANOVA techniques. Our objective is to uncover insights into the relationship between gender and visit time points on cognitive function, as well as to discern significant differences in brain volume across multiple visits for each patient. Furthermore, we aim to provide a comprehensive narrative that not only elucidates the story behind the data but also discusses potential avenues for further analysis to deepen our understanding of the research questions at hand.

2. Data Wrangling and Exploration

To address these questions, we commence by thoroughly examining the dataset, identifying relevant variables such as gender, visit time points, MMSE scores, and brain volume. While not every column may directly contribute to our analysis, we will meticulously explore the dataset to extract meaningful insights.

The dataset comprises 294 observations with 15 variables. The summary statistics reveal valuable insights into the distribution of key variables according to the research questions:

- MMSE (Mini-Mental State Examination): The scores ranging from 15 to 30.
- eTIV (Estimated Total Intracranial Volume): The mean eTIV is 1478.85, with a minimum of 1106 and a maximum of 2004.

We also check for missing values across all columns to ensure data completeness and integrity. This step is crucial for identifying any gaps or inconsistencies in the dataset that may affect our analysis. Since there is only 1 missing value the MMSE column, it is very clean, and there are 15 missing values in the SES column, which will not be

used in the research, so no further thing to do for dealing missing value.

Two key visualizations are provided to explore the relationship between visit time points (Visit) and cognitive function (MMSE scores) as well as brain volume (eTIV). The box plot depicting MMSE scores by gender and visit time points showcases potential trends or differences in cognitive function across visits and between genders. Additionally, the box plot for eTIV across visits highlights any variability or patterns in brain volume measurements over time.

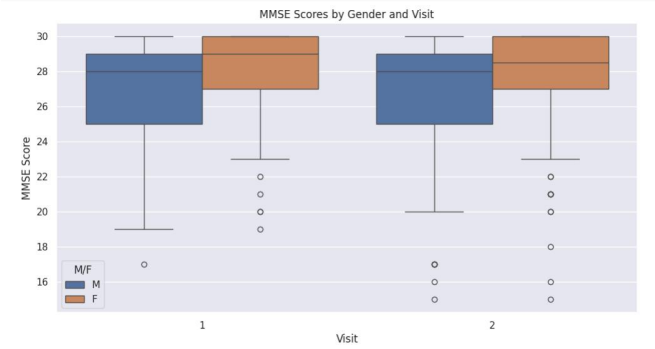

Figure 1: MMSE scores by gender and visit (question 1)

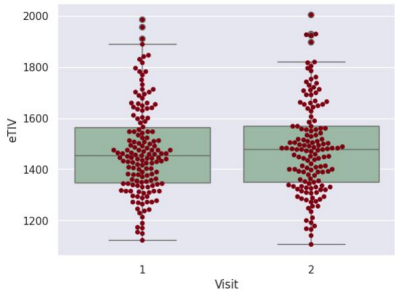

Figure 2: eTIV by visit (question 2)

These preliminary visualizations offer a glimpse into the dataset's characteristics and provide a foundation for more in-depth analysis using mixed-effects ANOVA. Further exploration and statistical analysis will be conducted to address the research questions and derive meaningful insights from the data.

3. Quantitative Analysis

3.1 Research Question #1: What is the impact of gender (M/F) and visit time points (Visit) on cognitive function, as measured by MMSE scores, while considering the interaction between gender and visit time points?

Upon conducting a mixed-design ANOVA to explore the impact of gender (M/F) and visit time points (Visit) on MMSE scores, several key findings emerged. Firstly, the analysis revealed significant main effects for both gender and visit time points. Specifically, the main effect of gender indicated a statistically significant difference in cognitive function between male and female patients ($F(1, 141) = 4.321$, $p = 0.039$), while the main effect of visit time points suggested variability in MMSE scores across different visits ($F(1, 141) = 8.610$, $p = 0.004$).

However, despite these significant main effects, the interaction between gender and visit time points did not emerge as statistically significant ($F(1, 141) = 1.606$, $p = 0.207$). This suggests that the influence of gender on MMSE scores remains consistent across different visit time points.

source	SS	DF1	DF2	MS	F	P-unc
M/F	88.69	1	141	88.69	4.32	0.039
visit	22.38	1	141	22.38	8.61	0.004
interaction	4.17	1	141	4.17	1.61	0.207

Table 1: Anova summary for impact of gender and visit on MMSE

Upon scrutinizing the assumptions of the ANOVA model, it became evident that the normality of residuals assumption was violated, as indicated by the Shapiro-Wilk tests for both Visit 1 ($p < 0.001$) and Visit 2 ($p < 0.001$). Additionally, the homogeneity of variances test (Levene's test) yielded inconclusive results due to missing values.

In summary, the analysis underscores the significant impact of both gender and visit time points on MMSE scores, highlighting gender-related differences in cognitive function and variations in MMSE scores over time. However, the consistent influence of gender across visit time points suggests a stable relationship between gender and cognitive function, warranting further investigation into the factors contributing to cognitive disparities between genders and changes in cognitive function over time.

3.2 Research Question #2: Is there a significant difference in brain volume (eTIV) at multiple visits for each patient?

The analysis aimed to investigate whether there exists a significant difference in brain volume (eTIV) across multiple visits for each patient. Visual inspection of the interaction effect between visit time points and eTIV revealed potential differences in brain volume across visits. Notably, there appeared to be fluctuations in eTIV over the course of multiple visits, suggesting variability in brain volume measurements.

Subsequent to visual inspection, a repeated measures ANOVA was conducted to assess the significance of the observed differences in brain volume across visits. The ANOVA revealed a significant main effect of visit time points on eTIV ($F(1, 143) = 9.247$, $p = 0.003$), indicating that brain volume varied significantly across different visits.

source	SS	DF	MS	F	P-unc
visit	5573.92	1	5573.92	9.25	0.0028
error	86200.58	143	602.80	NaN	NaN

Table 2: Anova summary for impact of visit on eTIV

Post hoc tests, employing paired t-tests with FDR-Benjamini-Hochberg corrections, were conducted to further elucidate specific pairwise differences in brain volume between visits. The post hoc analysis revealed a significant difference in eTIV between the first and second visits ($p = 0.003$), indicating that brain volume significantly changed between these two time points.

However, to ensure the validity of the ANOVA results, assumptions regarding the sphericity of the data were assessed using Mauchly's test. The test results indicated that the assumption of sphericity was violated for both visit time points (Visit 1: $p = 0.002$, Visit 2: $p = 0.016$), suggesting potential issues with the homogeneity of variance across visits.

In summary, the analysis revealed a significant difference in brain volume (eTIV) across multiple visits for each patient, as evidenced by the significant main effect of visit time points on eTIV. However, further investigation into the violation of sphericity assumptions is warranted to ensure the robustness of the ANOVA results.

3.3 Power Analysis

The power analysis aimed to determine the statistical power of t-tests conducted in the analysis. A priori power analysis was performed with predetermined parameters including an effect size of 0.7, a significance level (alpha) of 0.05, and a desired power of 0.91. The results of the power analysis indicated that a sample size of approximately 45.45 (≈46) would be required to achieve the desired power level under the specified conditions.

To visualize the relationship between sample size and statistical power, a plot of power against sample size was generated. The plot depicted an increasing trend in power with larger sample sizes, illustrating the impact of sample size on the ability to detect true effects. As sample size increased, statistical power approached the desired level of 1, indicating a greater likelihood of detecting true differences or effects when a larger sample size is utilized.

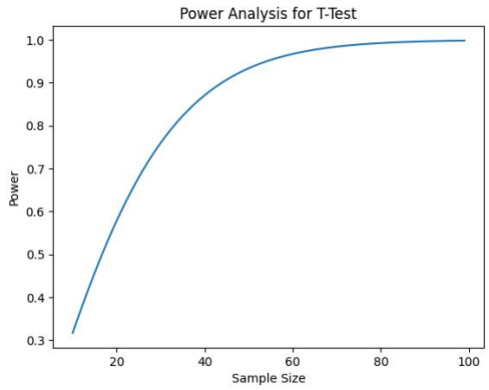

Figure 3: Power analysis

In summary, the power analysis provided insights into the relationship between sample size and statistical power, highlighting the importance of adequate sample sizes in ensuring the reliability and robustness of statistical findings.

4. Conclusion

In conclusion, research questions explore the impact of gender and visit time points on cognitive function, and differences in brain volume across multiple visits. Data exploration identifies key variables such as MMSE scores and eTIV measurements, with visualizations revealing trends over time. Quantitative analysis reveals significant effects of gender and visit time points on MMSE scores, though the interaction effect is insignificant. Besides, a significant difference in brain volume across visits is observed. A power analysis emphasizes the importance of sample size for reliable statistical findings. Overall, this study provides insights into cognitive health and brain morphology, suggesting avenues for further clinical research.